LA-UR- 10-00262

Approved for public release; distribution is unlimited.

Title: Uncertainty Quantification and Error Analysis

Author(s): Dave Higdon, Richard Klein, Mark Anderson, Mark Berliner, Curt Covey, Omar Ghattas, Carlo Graziani, Salman Habib, Mark Saeger, Joseph Sefcik, Philip Stark, James Stewart

Intended for: Workshop on Scientific Challenges in National Security: the Role of Computing at the Extreme Scale

Los Alamos NATIONAL LABORATORY EST.1943

Uncertainty Quantification and Error Analysis

Dave Higdon, Los Alamos National Laboratory
Richard Klein University of California, Berkeley and Lawrence Livermore National Laboratory
Mark Anderson, Los Alamos National Laboratory
Mark Berliner, Ohio State University
Curt Covey, Lawrence Livermore National Laboratory
Omar Ghattas, University of Texas
Carlo Graziani, The University of Chicago
Salman Habib, Los Alamos National Laboratory
Mark Seager, Lawrence Livermore National Laboratory
Joseph Sefcik, Lawrence Livermore National Laboratory
Philip Stark, University of California, Berkeley
James Stewart, Sandia National Laboratories

> The uncertainty is as important a part of the result as the estimate itself. … An estimate without a standard error is practically meaningless. H. Jeffreys (1967)

UQ studies all sources of error and uncertainty, including: systematic and stochastic measurement error; ignorance; limitations of theoretical models; limitations of numerical representations of those models; limitations on the accuracy and reliability of computations, approximations, and algorithms; and human error. A more precise definition for UQ is suggested below:

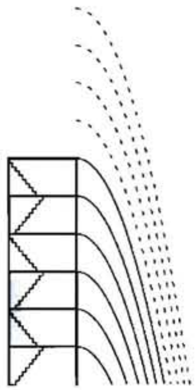
> Uncertainty Quantification is the end-to-end study of the reliability of scientific inferences.
>
> Ideally, UQ results in
> i. a quantitative assessment of that reliability,
> ii. an inventory of possible sources of error and uncertainty in the inferences and predictions,
> iii. an inventory of the sources of error and uncertainty accounted for in the assessment, and
> iv. an inventory of assumptions on which the assessment is based.

Uncertainty quantification (UQ) for estimation, prediction and assessment has long been held as fundamental to scientific investigations. Traditionally, UQ has been carried out via statistical analyses in applications ranging from drug efficacy trials to inferring the speed of light. Such analyses typically rely on a mix of theory, basic mathematical models, and sufficient observational or experimental data.

Advances in computing over the past few decades – both in availability and power – have led to an explosion in computational models available for simulating a wide variety of complex physical (and social) systems. These complex models – which may involve millions of lines of code, and require extreme computing resources – have led to numerous scientific discoveries and advances. This is because these models allow simulation of physical processes in environments and conditions that are difficult, or even impossible, to access experimentally. However our ability to quantify uncertainties in these model-based predictions lags well behind our ability to produce these computational models. This is largely because such simulation-based scientific investigations present a set of challenges which are not present in traditional investigations:

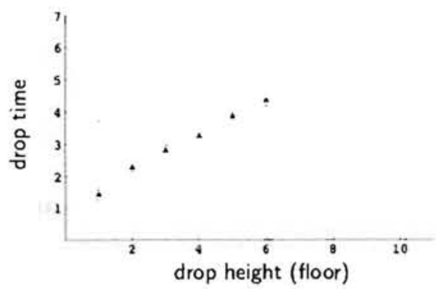
- The amount of physical data (observational or experimental) is typically quite limited;
- The computational demands of the model limit the number of simulations that can be carried out;
- The computational models are not perfect representations of physical reality – they have inadequacies, approximations, missing physics, etc.;
- The computational models typically have unknown parameters and boundary conditions which need to be adjusted for the application at hand;
- We often wish to use such models in extrapolative conditions, where we have little or no physical observations to validate model output.

Sidebar 1 gives a very simple example of how experimental observations and a computational model are combined to infer times of flight for an object being dropped from a tower. In addition to leading to more accurate predictions, advances in UQ methodology will lead to more reliable uncertainties for simulation-based predictions. This is particularly important in high consequence decisions for which both under and over stating uncertainties leads to excess costs or liabilities. More importantly, new UQ methods for simulation-based investigations will lead to improved understanding of the different sources of uncertainty affecting predictions. This will allow decision makers to be more effective with their limited resources. For example, one can ask how can I best use my available resources to reduce uncertainties? Should I improve computing resources? Carry out new experiments (which ones)? Improve experimental diagnostics? Improve existing computational models?

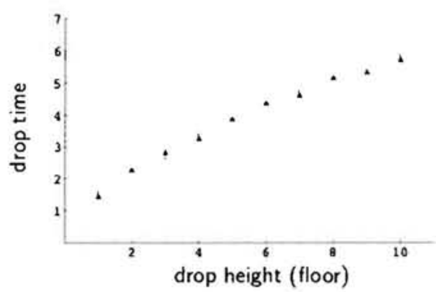


A simple UQ example:
Using experimental data and a computational model to predict drop times from new heights.

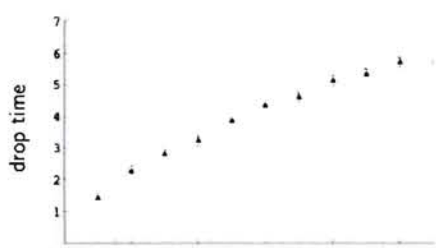
Experimental data:
The time it takes an object to drop from each of 6 floors of a tower is recoreded. There is an uncertainty in the measured drop times of about $\pm .2$ seconds. Predictions for times are desired for drops from floors 7 through 10 – which do not (yet) exist.



Simulated drop times:
A (trivial) computational model is developed to predict the drop times as a function of drop height. The simulated drop times (red line) are systematically too low when compared to the experimental data (plotting symbols). The error bars around the observed drop times show the observation uncertainty.



Accounting for modeling error:
This systematic discrepancy between the computational model and the experimental data is accounted for with a statistical adjustment. This term adjusts the model-based predictions to better match the data. The resulting 90% prediction intervals for floors 7 through 10 are also shown in this figure. Note the prediction intervals become wider as the drop level moves away from the floors with experimental data. The points corresponding to floors 7 through 10 show experimental observations taken later, solely for evaluation of the predictions.



Improved mathematical model:
An improved simulation model was constructed which accounts for air resistance. A parameter controlling the strength of the resistance must be estimated from the data, resulting in some prediction uncertainty (90% prediction intervals are shown for floors 7 through 10). The improved model better captures the physics in the process, giving reduced prediction uncertainty.

Current Status

Presently there is a substantial amount of research activity devoted to inference and uncertainty quantification aided by computational models. Below we outline some of the main areas of current research. These current research areas relate to, and help motivate the principle research directions presented at the end of this section.

Inverse Problems and Calibration of Computational Models

It is often the case that a computational model requires physical observations to adjust key model parameters, initial conditions and/or boundary conditions to better model the physical system. In a typical inverse problem, these quantities are determined by minimizing the discrepancy between physical observations and computational model output. Statistical approaches to inverse and calibration problems require that this discrepancy between observations can be formalized into a likelihood function which is produced from a probability model for the data given the model parameters. More formal statistical inference about the unknown parameters and/or initial and boundary conditions can then be made, describing their uncertainty. Because many of these inverse problems are ill-posed – especially when estimating a large field of initial conditions – it is necessary to regularize (enforce "smoothness" or other properties) on the unknowns being estimated. Bayesian methods for statistical inversion and calibration have become popular since these approaches codify regularization in prior distributions and give a probabilistic description of the resulting uncertainty. However, interpreting these resulting probabilities can have its own set of difficulties. In practice, solving problems in inversion and computer model calibration can be complicated by a variety of issues, including high-dimensional parameter spaces, computationally demanding forward models, nonlinearity and/or complexity in the forward model, sparse physical observations, and inadequacies (numerical and physical) in the computational model.

Sensitivity Analysis

Sensitivity analysis is the systematic study of how model inputs – parameters, initial and boundary conditions – affect key model outputs. Depending on the application, one might use local derivatives or global descriptors such as Sobol's functional decomposition or variance decomposition. Also, the needs of the application may range from simple ranking of the importance of inputs to a response surface model which predicts the output given the input settings. Such sensitivity studies are complicated by a number of factors, including the dimensionality of the input space; the complexity of computational model, limited forward model runs due to the computational demands of the model, the availability of adjoint solvers or derivative information, stochastic simulation output, and high-dimensional output. Challenges in sensitivity analysis include dealing with these factors while addressing the needs of the application.

Predictions from Multi-model Ensembles

In applications such as climate and weather prediction, a variety of different models are available for making a particular forecast. Predictions and estimates of uncertainty that combine the results of multiple models often give better predictions than any single model. Recent research has focused on approaches for combining results from different models. This includes approaches based on Bayesian model averaging as well as game theoretical paradigms. Weather forecasts based on such approaches have proven very successful (see the University of Washington's Probcast, or the Canadian Weather Office's ensemble forecasts, for example).

Representing Uncertainty

The question of how to represent and communicate uncertainties is a topic of research both from a practical and theoretical point of view. A fair bit of theoretical research is aimed at the mathematical calculus of uncertainty. This includes extensions and alternatives to standard probabilistic reasoning, such as Dempster-Schafer theory and imprecise probabilities. When uncertainties are needed for investigations requiring computational models, additional considerations arise. For example, if the simulation output is a daily surface temperature field over the globe for the next 200 years, representing uncertainty and dependencies is complex. Should ensembles be used to represent plausible outcomes? How should these ensembles of simulation output be stored? How can high consequence-low probability outcomes be discovered in this massive output? Here some research investigations look to leverage theory that exploits high-dimensionality to bound probabilities and system behavior. Finally, even when uncertainties are well captured, how best to communicate such uncertainties to the public, or to decision makers is also a topic of ongoing research.

Verification and Validation

Verification and validation (V&V) has been a staple of the computational model assessment community for the past few decades. Standard definitions are given below.

> **Verification** is the process of determining as completely as possible: 1) whether a computer code correctly implements the intended algorithms; and 2) the accuracy with which the algorithms solve the intended equations.
>
> **Validation** is an assessment of the degree to which predictions of a code represent the intended physical phenomena, with the purpose of quantifying how accurately the model equations represent physical reality over a specified regime of applicability.

The two V's in V&V have focused on the comparisons: code vs. math model; and code vs. reality. Clearly both of these activities are related to UQ. Research in verification includes convergence assessment, estimation of bias and uncertainty due to numerical error, a posteriori error estimation, and the method of manufactured solutions. Validation focuses on comparing simulation-based predictions to experimental results. Research in validation includes choosing/designing experiments, assessing experimental uncertainties, propagation of uncertainty and determining the physical regime in which a code is validated.

Clearly, much of the activity in validation overlaps with UQ and much of the research from this area is relevant to UQ. In many settings V&V has a regulatory flavor. Hence V&V may not focus on other questions that are clearly in the realm of UQ. For example: How can a model be adjusted to give improved predictive accuracy? How should different sources of uncertainty be combined to determine prediction uncertainty? How should one produce uncertainties for predictions outside of the validation regime?

Data Assimilation

A number of applications in monitoring and surveillance require persistent updating of the state of the system, predictions and uncertainties based on continual or periodic collection of new physical observations. This data is combined with the computational model to update inferences. When the physical system is linear, and observation and model evolution errors are Gaussian, this updating can be accomplished with the Kalman Filter, which updates the state of the physical system in an iterative fashion using the new data. More recent research has focused on updating large-scale, non-linear systems – such as interacting particle systems, oceans and atmospheres. Here the nonlinearities, massive data (e.g. satellite and sensor readings), and the extreme computational effort required to run these models have made it necessary to develop new approaches for updating information about the physical system and producing predictions with uncertainties. This has motivated much recent research effort on extensions of the Kalman filter, such as the ensemble Kalman filter, extended Kalman filter, and Monte Carlo based techniques such as the particle filter.

Optimization, Adaptive Design and Feedback

These large-scale computational models are a powerful tool for planning and decision-making. Such models can be used to help assess important questions regarding the management of a particular system. For example, how many and what types of sensors are required to monitor the planet's greenhouse gasses over time? How should limited resources be used for this monitoring? Airborne measurements? Remote sensing? Ground-based sensors? Also, what are the likely responses to different mitigation actions? These management questions – which have analogs in almost any other application – require understanding of uncertainties, and how various actions will reduce, or otherwise affect, these uncertainties. Research in this area includes optimization, decision theory, the design of experiments and optimal sensor networks, and optimal resource allocation. This line of research typically comes with a much larger computational burden since the impact of many possible (mitigation, allocation, design, ...) strategies must be assessed in order to find one that is optimal – or near optimal. The difficulty of such problems is further complicated by the computational demands of the models, as well as by the fact that the model may deviate from reality in ways relevant to addressing the optimization question.

Basic Science Challenges and Research Needs

The Role of Uncertainty Quantification in Extreme Computing

Extreme scale systems present daunting challenges for their usage. Considering these challenges in the context of a UQ workload we are presented with unique opportunities to think about these challenges differently than previous terascale and petascale systems and thus make these problems more amenable to solution.

The DOE Exascale Initiative envisions the development and deployment of Exascale systems in the 2018 timeframe. These systems will bring unprecedented computational power to pressing scientific simulation activities for both Office of Science and NNSA. However, the technological trends underpinning these systems will push the architectures in directions that will present many challenges to their gainful employment for large scale predictive scientific simulation based scientific discovery activities.

Current computer industry trends portend significant challenges to scientific simulation in general. New programming models will probably be required in the Exascale generation of platforms to deal with the O(1B) way parallelism and also application resiliency challenges. Also, new usage model paradigms need to be considered. This is where UQ might make a very important contribution to our usage model for Exascale systems. For example, today applications are considered as a stand-alone single job running across the entire system or space sharing the system with a few other large applications. This requires a very high Mean-Time Between Failure (MTBF) for the hardware and software, because any non-redundant hardware failures cause the application to abnormally terminate. The most common way of dealing with the fact that the MTBF of today's systems (days to a week) is usually much shorter than the mean run-time of those applications (weeks to months) is for the application to periodically checkpoint the internal state of the calculation and restart from the most recent checkpoint upon abnormal termination. However, the cloud computing model used for most Web 2.0 services (e.g., Amazon, Google, Yahoo!, eBay, etc) use a transaction model of service where each HTTP request can time-out and be retried. If there is a hardware or software failure of a component part causing a HTTP request to fail, another host can seamlessly take over servicing the retry. That way the overall service can have 99.999% availability but the MTBF of the underlying hardware can be fairly low. Indeed, Google is very proud of the "extremely inexpensive" hardware they use for their cloud computing infrastructure. Because future UQ "throughput" workloads of O(10K) to O(100K) ensemble jobs may be typical, one can think of each job or instantiation of the predictive simulation application running on the Exascale platform as a "transaction" that can be retried (or transparently restarted from a checkpoint) upon failure. This approach completely changes the application resiliency problem that must be solved for Exascale systems.

Another challenge for UQ ensembles on Exascale systems (and petascale systems before them) is the vast quantity of data that is generated during the ensemble runs. It is not just the amount of data (measured in 1s-10s of ExaBytes or EB) that is challenging, but the vast quantity of files and directories used to map data back to an individual run. Without a database keeping track of this mapping, it will be impossible to manage the Ensemble data. Most modern codes are written in object oriented (OO) languages (e.g., python, C++) in an OO style. Most UQ Frameworks or "pipelines" are also written with OO languages and techniques, but those object hierarchies are separate from those of the applications they manage. Also, several parallel file systems are object oriented and utilize object storage devices (RAID devices or the disks themselves). However, the file system Objects do not map onto the storage device objects. Thus, key information about the data to be stored (metadata) is lost.

One of many possible approaches to address these issues is to adopt a new file system paradigm that allows the UQ pipeline to define the objects that are used and augmented by the applications it drives and the packages within those applications can augment these objects and then pass them directly to the file system with the full context of the computation and IO operation (e.g., on timestep X for package A within Application B working on parameter study Z for UQ study alpha) directly into the file system. Then the file system becomes an object oriented database that allows one to search on application defined metadata parameters consistent with the object oriented hierarchy.

Additional challenges for UQ on Exascale platforms, due to their vast scale, will be job management (e.g., job preparation, execution, statusing, results analysis and termination). The vast scale of Exascale platforms will make this problem more difficult because of the quantity of jobs being run and the need to deal with a necessarily large number of things that could (and will) go wrong. A UQ capacity workload interspersed by "full system" UQ jobs will really stress future job scheduling and allocations management infrastructure. However, the UQ capacity workload(s) could play an important "bottom feeder" role by soaking up otherwise idle cycles and by presenting a set of application transactions that can be terminated easily when "full system" runs need to be launched and then immediately fill in the gaps caused by abnormal termination of the large job(s). In fact, if the scheduler could communicate with the UQ framework or pipeline and request jobs of various sizes with specific runtimes on demand, based on the current system load and expected future load based on the queued jobs, then the scheduler could predictably fill empty run/time slots on the machine. This flexible runtime model could be used to keep Exascale system utilizations very near 100%.

The above three examples show that thinking about Exascale systems in the context of a UQ workload and in conjunction with a "full system" workload, provides many opportunities to fundamentally change the problems that need to be solved to make Exascale systems deliver predictive simulation results with confidence. Some of these changes of the problem to be solved make it much easier to actually solve them than previous approaches for terascale and petascale systems with static large job workloads. In other words, UQ changes the game for Exascale systems and vice versa.

The above discussion has focused on the "UQ as Throughput" system usage model. Here a sequence of forward model runs – at various initial conditions and parameter settings – are used to understand sensitivities, propagate uncertainties, constrain parameter uncertainties, make predictions and estimate probabilities. Figure 1 shows how a response surface can be used to interpolate model output and constrain parameter uncertainties. Depending on the problem, UQ as throughput will require from 10^3-10^8 computational model runs. These runs tend to be large in quantity and generate an ExasFLOP/s computing load in aggregate. They also tend to generate vast quantities of data that need to be stored on a shared parallel file system and later analyzed as a single system or a tightly coupled simulation environment with a global parallel file system and low latency, high bandwidth Storage Area Networks (SAN).

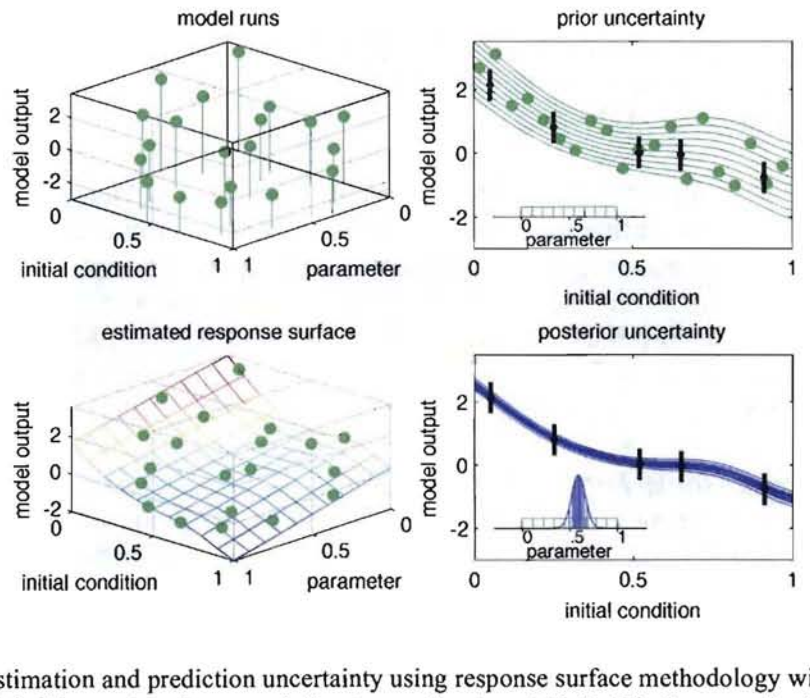


Figure 1: Parameter estimation and prediction uncertainty using response surface methodology with limited model runs. Top left: Model runs have been carried out according to a statistical design over the input settings determined by the initial condition and physical model parameter. The computational model output is given by the circle plotting symbols. Top right: The prior density for unknown model parameter and the implied simulations using quantiles of the prior parameter uncertainties are given by the green lines. The black dots show physical observations and the corresponding black lines give 95\% uncertainty bounds on the observations. UQ uses both the data points and the model runs to reduce uncertainties in the parameter and in the prediction. Bottom left: The prior uncertainty for the computational model prediction as a function of initial condition; the actual model runs are marked by the circle plotting symbols. Bottom right: The darker lines

show the updated uncertainty for the model parameters and the resulting model-based predictions. Uncertainty in the model prediction is due to uncertainty regarding the parameter as well as uncertainty in the response surface estimate shown in the top right frame.

On the other side of the spectrum is the challenge of carrying out UQ on computational models that require extreme computing for just a single simulation. Here one can expect only a handful of high fidelity forward runs with which to carry out UQ. While this is impossible with most computational models, this may be possible if the next generation of computational models are constructed with UQ in mind. For example computational models may be equipped with options for running faster, more reduced models which give more approximate results. Another possibility is to augment computational models with adjoint solvers that can be used to compute derivatives of important outputs with respect to key inputs. In any case, for problems that require extreme computing resources for a single run, UQ methods need to be co-designed with the computational model to allow exploration of sensitivities and uncertainties. Clearly, this is a research topic that should accompany the development of computational models for applications deserving of extreme computational resources.

A Key Motivating Application for Uncertainty Quantification: Climate

An example of a scientific Grand Challenge problem requiring a UQ analaysis is the prediction of the future climate. The most important factor in climate prediction is Earth's equilibrium climate sensitivity (Bader et al. 2008). By definition, this is the increase in globally averaged surface temperature that would result if atmospheric carbon dioxide doubled, all other climate-forcing agents remained the same, and enough time elapsed for a new statistical steady state or "equilibrium" climate to be established. Equilibrium climate sensitivity may be thought of as climate response per unit of climate forcing. It is important to practical applications of climate prediction because local climate impacts generally scale with globally averaged temperature change (Santer et al. 1990).

Arrhenius in 1897 concluded that Earth's equilibrium climate sensitivity is roughly 3 degrees C. That estimate changed very little during the ensuing century. The US National Academy of Sciences (Charney et al. 1979) concluded that equilibrium climate sensitivity probably lies in the range 1.5 – 4.5 degrees C, a statement repeated in subsequent Intergovernmental Panel on Climate Change (IPCC) assessment reports (e.g. McAvaney et al. 2001). In the late 1990s, work in the United Kingdom began that substantially advanced the state of the art of climate-uncertainty quantification. This work employed large (~104) ensembles of climate model runs with differing input assumptions and produced, for the first time, probability density functions (PDFs) of equilibrium climate sensitivity (Meehl et al. 2007).

The PDFs are all consistent with the earlier, more qualitative estimates of equilibrium climate sensitivity cited above. Different and apparently equally sound methods, however, give significantly different PDFs. Furthermore, all PDFs produced to date are very broad, and as a result the wide uncertainty range estimated over thirty years ago by the National Academy of Sciences has not been narrowed. In fact, higher sensitivity values (> 5 degrees C) have non-negligible probability according to most of the PDFs, and at the other end of the range, values < 1 degree C are implied by some calculations (e.g. Lindzen and Choi 2009).

Despite this rather unsatisfactory state of affairs, detailed probabilistic climate forecasts are starting to appear, beginning with the UK Climate Projections issued in 2009 by the United

Kingdom Meteorological Office (http://ukclimateprojections.defra.gov.uk). Clearly a reasonable path forward must include sounder scientific underpinning for such climate predictions, which will be forced "by popular demand" whether or not scientists think they are ready to provide them. Therefore we believe that uncertainty quantification will play a high priority role in climate science.

Comprehensive UQ studies in the climate domain requires methodology to cope with high-dimensional uncertain input space and methodology to compare and contrast high-dimensional model output to observation-based datasets of various quality. UQ studies would yield understanding about the relationship between the uncertain climate processes, benefiting future climate-model development. UQ studies would also yield practical assessment of climate projection uncertainty, benefiting current climate-impact assessments. Propagating uncertainty in global climate model projections to regional scales introduces additional challenges; variation in regional fidelity (e.g., biases) between global climate model projections which are all judged to be equal at the global scale. This applies particularly to regional precipitation. However, such regional UQ is necessary and crucial for practical climate-change impact studies and assessment of different climate-policy options.

Priority Research Directions

Uncertainty Quantification is an emerging field that presents many research challenges. These challenges will best be addressed by a focused effort to identify priority research directions that we believe are opportunities to further advance the field. While UQ is a rather general topic, with many possible research directions, we focus here on research directions that are of particular importance in light of extreme scale computing.

Foundations in UQ

Everyone has an intuitive notion of uncertainty as doubt or a lack of certainty regarding a possible outcome of an anticipated event. This is particularly important to decision makers in national security applications who must assess the chances of numerous adverse events and take action to protect us from the most dangerous of these, with limited resources. For repeatable events we can assess uncertainty statements by comparing them against actual results. For example, we can check to see how often alleged 90% intervals for tomorrow's temperature actually contain that day's temperature. For well-calibrated UQ methods, the 90% intervals should contain the actual temperature 90% of the time. UQ methods that make better use of available information and computational models can produce much tighter intervals that still cover 90% of the time – see Figure 2.

While this type of assessment is quite intuitive and sensible, assessing predictions and uncertainties for rare, or "one of a kind" events is much more problematic. For example, what is the chance of a substantial terrorist attack in a major US city? What is the chance the ocean level rises by three meters within the next 40 years? What is the chance of a sizable asteroid impacting the earth in the next 1000 years? Assessing such uncertainties is difficult; nonetheless, decision makers still must decide what resources and actions will be devoted to mitigating these types of risks. An assessment of the uncertainties associated with such events is fundamental to these decisions.

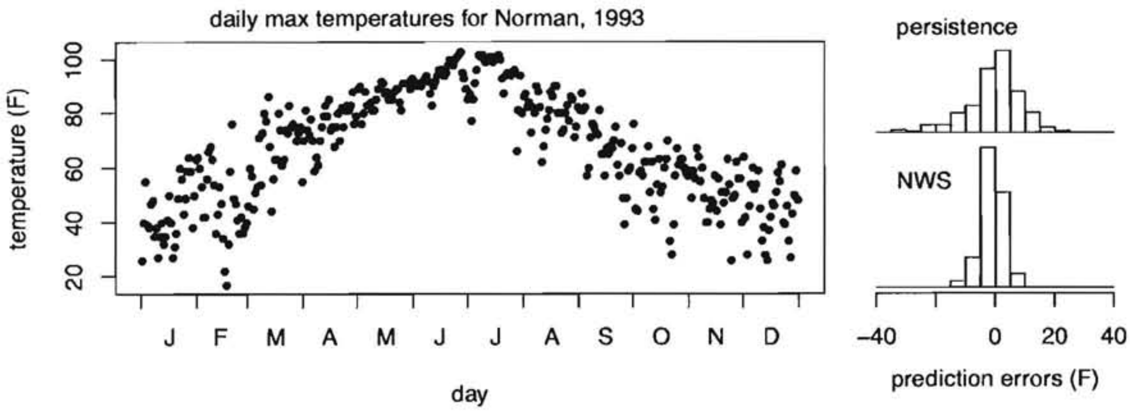


Figure 2 One day ahead predictions for the daily maximum temperature for Norman OK using two forecasting models: persistence – predict tomorrow's temperature with today's temperature; and the National Weather Service (NWS) forecast. 90% of the actual temperatures are within ±14°F for the persistence forecasts, and ±6°F for the NWS forecasts. The greater accuracy of the NWS forecasts is due to its use of computational models and additional meteorological information. The assessment of these two forecast methods is straightforward because of the large amount of replication.

Scientific and Computational Challenges

We expect that large-scale computational models, which encode numerous physical laws and phenomena, are viable tools to help with this uncertainty assessment. These models are clearly more than empirical, statistical formulations. We need strong foundational underpinnings to make the best use of these computational models for important decisions regarding difficult to assess events. This PRD focuses on the development of language and framework for linking these promising computational models to reality, producing meaningful statements of uncertainty.

The basic notion that better models give better predictions has not been formalized or mathematically encoded for assessing uncertainties in predictions from computational models. Current approaches, such as those in statistics and machine learning, make no theoretical distinction between a physically motivated computational model and an empirically based regression model – the better model eventually proves itself on physical observations that have been held back to test the competing predictions. The challenge is to develop a framework that better accounts for the nature of the model being used to make predictions, leading to meaningful prediction uncertainties. This framework should be helpful even when the physical data available to train and assess these models are limited.

Summary of Research Direction

The issues with uncertainty quantification are present at all scales of scientific computation form the desktop to the exascale, and include both descriptive and quantitative elements. A consistent treatment of uncertainty requires a consistent semantic basis, a language in which concepts regarding uncertainties in the context of physically-based, computational models can be communicated.

In addition to the semantic and mathematical fundamentals, a disciplined process for performing, documenting and assessing uncertainty quantification is required to support decision makers. This process must include the universal glossary of terminology and standard guidelines for analysis and documentation, including descriptions of the problem at hand, assumptions made, and methods applied. Efficient application of a standardized process is facilitated by examples of its application to relevant problems.

Expected Computational and Scientific Outcomes

The products of this research direction are things needed to perform uncertainty quantification coherently and make the results meaningful and useful to decision makers. These products include the following:

- Methods for assessing the usefulness and quality of uncertainty quantification analyses;
- Methods for analyzing the sensitivity of uncertainty quantification to the assumptions and methodology;
- Methods for assessing uncertainties of model-based predictions in new, untested regimes (i.e. "extrapolations");
- Methods that leverage theoretical considerations and computational models to assess the likelihood of extreme, high consequence events;
- Reporting guidelines for estimates and uncertainties, including disclosure of assumptions and methods;
- Compelling examples of uncertainty quantification done well in problems with different degrees of complexity.

Potential Impact on National Security

Achievement of the research goals will result in consistent, quantified support for decision makers. This research will also lead to uncertainty assessments with higher quality and more common features. The variety of applications of such support covers the entire spectrum of activities supported by simulation at any scale up to and including the exascale. This research should help illuminate when high fidelity, first principles models are required, and when they are not. Examples of potential applications include stockpile stewardship, nuclear reactor safety and many other aspects of the nuclear security enterprise, considerations of legislative and judicial options for actions to address climate change, and the use of very large scale computing as a surrogate for experiments of discovery in basic science.

Countering the curse of dimensionality

Scientific and Computational Challenges

The current state-of-the-art for UQ science of multi-physics simulation codes is done with an ensemble of models approach. That is, we take a model of interest and we identify the subset of input quantities (usually ~7-10) that we think will dominate the uncertainties in the predictions of interest since we have limited computational power. We then compute the model thousands of times with differing combinations of input quantities (ensembles) and we constrain some or all of the inputs by using available data with their own associated uncertainties. Finally we compute bounding uncertainties for predicted model outputs. Unfortunately, there are many limitations to our current capability and in particular, one stands out. To provide more accurate uncertainty bounds, we need to be able to vary all of the input quantities that have uncertainties that can influence a given prediction. Generally, we know that of order 100 input parameters (e.g. Climate, ICF) in many multi-physics areas can influence simulation predictions. Thus, to advance UQ science we need to develop new methodologies that can accommodate the vast number of existing uncertainties (dimensions) with a minimum of code calculations. This challenge is referred to as the Curse of Dimensionality. To do so in an efficient manner, we need an automated process (UQ Pipeline) that can self-guide the large numbers of ensemble simulations needed to explore the complexity of having uncertainties in many 10s to 100s of input parameters and which can identify and apply the most efficient combination of methodologies for minimizing the number of code runs required. Such a UQ Pipeline will enable us to test and refine the UQ methodologies as they are developed and it will eventually enable efficient utilization of Exascale computing. Current practice to reduce the numbers of simulations required is to constrain parameter ranges (i.e. input space) based on available observational data, physical considerations, and/or the results of previous studies. Two simple and commonly used approaches to constructing the ensembles are Monte-Carlo and Latin Hypercube sampling of the input space. Response surfaces (also known as a statistical response model, a surrogate model, a meta model) are then constructed from the ensemble results and these are then convolved with observational data to further constrain input parameters and to create uncertainty bounds on model outputs. This approach works well for studies limited to the variation of a handful of input parameters. However, this does not hold when the simulator is computationally expensive and when the uncertain input-parameter space is high-dimensional. This is the case for comprehensive UQ for multi-physics, multi-scale codes (e.g. global climate models).

Since the problems we are interested in have uncertainties associated with 10s to hundreds of input parameters, we are saddled with the "curse of high dimensionality." For computationally expensive simulators and current sampling methods, this makes comprehensive UQ intractable with current and even future envisioned computational platforms. Put simply, Monte Carlo and Latin Hypercube sampling techniques cannot adequately resolve the output volume space when the input space is composed of 10s of dimensions. For high-dimensional input spaces, it is near impossible to design a computer experiment up front that captures all the important aspect of the input-output relationship, which has motivated various adaptive (sequential) sampling strategies.

Summary of Research Direction

There is a need to reduce the effective size of the input-space, either through formal dimension reduction techniques or through input variable selection methods. Research in several areas may be promising.

Self-adaptive exploration of response surfaces: Methods to produce an efficient and robust ensemble of simulation results through adaptive sample refinement (ASR) need development. These methods may guide the ASR process through topological characterization and also through deficiencies in the predictive accuracy of the response model. Level sets provide iso-parametric contours of the response function. When the model is high-dimensional, finding the boundary of such sets accurately is extremely difficult methods need to be developed to refine adaptively near the boundary of the set.

Large-scale parallel response surface analysis methods: Global response surface approaches to uncertainty quantification – using the "ensemble of models" method need to be broadened. Generalization and extension of regression models to include basis functions with local support for the purpose of enhancing an ASR capability need to be researched as well as parallel, scalable algorithms for the purpose of generating response functions of very high dimensionality.

Theory/methods for high dimensional representations: Contour tree approaches can be developed to analyze the topology of high-dimensional functions. This can be used to improve ASR by identifying monotone/non-monotone regions to guide sampling of high dimensional response surfaces. New approaches would include basis function transformations. Specifically, the objective is to transform the original high-dimensional parameter space into a lower pseudo-dimensional space that can be used as a surrogate for the original input space to reveal underlying structure in the original space.

Surfing the UQ pipeline: Future Exascale studies will likely consist of tens of thousands or greater numbers of ensemble simulations and the sample space studied will be composed of a myriad of uncertainty dimensions and complexity. This class of UQ studies will be intractable if the user is required to make decisions concerning the guidance of the UQ study. This complexity requires a UQ Pipeline to include self-guiding, self-adapting technologies that steer the ensemble of simulations, without user interaction, toward the areas in the sample space where the effort should be focused.

Expected Computational and Scientific Outcomes

With the development of advanced methodologies to attack the curse of dimensionality, it will become possible to perform UQ analysis across a broad range of multi-physics, multi-scale scientific problems that include the main uncertainties inherent in the underlying physics models, numerical algorithms, data bases, inputs and output observables. Such developments will also result in a self-adapting, self-guiding UQ pipeline that will enable UQ studies to be performed on Exascale platforms.

Potential Impact on National Security

Progress in key areas of UQ research such as the curse of dimensionality will impact critical areas of importance to national security such as nuclear weapons and stockpile stewardship science; climate prediction and inertial confinement fusion (ICF) to name a few. In the area of climate prediction, it will likely be possible to make consistent uncertainty estimates in global climate sensitivity; predict regional climate impacts and move to Exascale computing within 8 years to include vastly improved cloud physics.

Intrusive/embedded UQ

Scientific and Computational Challenges

There are three fundamental components to end-to-end UQ for large-scale simulations (whether in the form of PDEs or ODEs or integral equations or discrete particle systems or other simulation models): (1) the statistical inverse problem: estimation of uncertainty in model parameters or model structure from observations or measurements; (2) the uncertainty propagation problem: propagation of input parameter uncertainties through the simulation model to predict model outputs; and (3) the stochastic optimization problem: solution of optimal design or control problems that are governed by the stochastic forward problem and that make use of statistics of model predictions as objectives and/or constraints. Unfortunately, contemporary techniques for solving the stochastic inverse, forward, and optimization problems suffer from the curse of dimensionality, and become computationally intractable for problems governed by large-scale simulation models with high-dimensional uncertainties. The availability of exascale computing, by itself, will not overcome these challenges; we need fundamentally new algorithms and analysis for estimation, propagation, and optimization under the presence of uncertainties in large-scale simulations of complex systems.

Summary of Research Direction

We believe that one of the keys to overcoming the twin curses of high-dimensionality and expensive forward simulations in UQ methods is to exploit the structure of the mathematical model that maps parameter inputs to output quantities of interest. Most contemporary UQ methods, such as conventional Monte Carlo methods, treat this input-output map (i.e. response surface) as a black box. Yet recently developed methods that exploit this input-output map structure have been critical for solution of deterministic inverse and other optimization problems with millions of parameters at a cost of a handful of forward simulations, and for greatly reducing the cost of sampling by approximating response surfaces and constructing reduced order models as surrogates for expensive forward simulations. More recently, similar ideas to exploit the mathematical structure of the input-output map have begun to appear in UQ methods for forward and inverse propagation of uncertainty. Many of these methods employ rapidly computed derivative information, motivated by the fact that for most systems governed by differential (and related) equations, the outputs are locally smooth and thus derivative information is generally useful. We refer to such methods as "intrusive" or "embedded" methods, since they require access to and analysis of at least the Jacobians of the underlying forward operators. Several examples of new research directions in UQ based on intrusive methods are given below.

Langevin methods for sampling probability densities, whose trajectories are driven by gradients of the (log of the) target density, are beginning to be employed for high-dimensional PDE-based inverse problems. Computation of the gradient is greatly facilitated by the use of adjoint equations (whose operator is the adjoint, or transpose, or the linearized forward operator). The gradient can be computed at a cost of at most a single forward simulation, and usually less, since the adjoint equation is always linear even when the forward problem is nonlinear. The price one pays for this capability is that legacy forward simulation codes often have not been designed to compute adjoints, and retrofitting complex legacy codes with adjoint capabilities entails a significant refactoring. Hessians (i.e. second order sensitivities) provide even richer information than gradients, and play a critical role in identifying significant directions in high-dimensional inverse problems, i.e. those directions in parameter space for which the data provide meaningful information on the model. Moreover, recent techniques to build low rank approximations of data-misfit Hessians based on analysis of the underlying infinite-dimensional operators have allowed for significant acceleration of sampling methods in inverse problems.

Hessians are also beginning to be used in the construction of reduced order models as surrogates for expensive forward simulations, where they can efficiently steer the placement of design points in parameter space at which full-order model outputs are computed and employed for reduced model construction. Low-dimensional reduced models constructed in this way have proven to be very effective at approximating full-order model outputs, and in turn have facilitated rapid sampling of probability densities that embed expensive forward simulation. Similarly, Hessians have begun to be used in the construction of Gaussian process approximation of response surfaces, both in effectively placing design points in parameter space as well as in informing the Gauss process approximation.

Polynomial chaos methods are another technique for exploiting the mathematical structure of the parameter-to-output map, in this case by approximation by multivariate polynomials. While convergence can be very fast, standard PC methods suffer from the curse of dimensionality and they soon become unaffordable. Sparse grid approximation techniques improve the convergence of PC methods, and even greater improvement can be obtained using anisotropic sparse approximation, which exploits the relative importance of different input random variables on the solution.

Still, despite very promising performance exhibited by emerging intrusive UQ methods for some challenging problems, it is fair to say that such methods are in their infancy, and substantial works lies ahead in extending, robustifying, tailoring, and scaling up these methods to address the complex, large-scale, non-linear multiphysics, multiscale, high-parameter-dimension UQ problems arising in national security applications. To do this, fundamentally new ideas are needed to further exploit mathematically and computationally the functional relationship between input parameters and output quantities of interest.

Expected Computational and Scientific Outcomes

As discussed above, current methods for quantifying uncertainties in simulation models are incapable of scaling up to expensive simulations characterized by large numbers of uncertain parameters. The development of UQ algorithms and methods that can exploit input-output structure to overcome these pervasive and abiding barriers will enable computational scientists to carry out large-scale simulations with quantified uncertainties, thereby transforming our ability to effect meaningful predictions for many critical scientific, societal, and strategic problems.

Potential Impact on National Security

One of the central challenges facing the field of national security is to employ large-scale simulation as a tool for decision-making involving uncertain complex systems. For such problems, the "single point" deterministic predictions produced by contemporary large-scale simulations are of little use for decision-making: to be useful, these predictions must be accompanied by estimates of their uncertainty. Many problems in the national security portfolio are characterized by large-scale, expensive simulations and high-dimensional parameter spaces. Intrusive UQ methods promise significant breakthroughs in our ability to address high-dimensional uncertainty and expensive simulation models. Success in developing such techniques will lead ultimately to a revolution in the way that decision-making under uncertainty is conducted, by permitting predictive simulations to be employed in a tightly-integrated way.

UQ in data-rich environments

The role of data, in terms of size, quality, and access, is becoming ever more prominent in all areas of human activity. The dramatic changes underlying the data revolution have been brought about by rapid progress in solid-state technology, ubiquitous networking and sensing, cheap storage, and associated advances in the computer and information sciences. While no one doubts the many uses of the analysis of datasets that previously could not even be imagined, one must deal with handling the "data flood" – a runaway generation of data that can only be contained by a corresponding exponential increase in our ability to ingest, store, organize, and interrogate the datastream. From the analysis perspective, it is useful to think about two extreme classes of problems, those related to real-time (or near real-time) applications and those related to dealing with very large datasets – both observed and simulated. In both cases, uncertainty quantification must play an essential role in confronting the data flood.

Scientific and Computational Challenges

There are a significant number of scientific and computational challenges posed by large datasets and data throughput. Areas in which these arise include astrophysics, biology, climate modeling, cyber security, earth sciences, nuclear and particle physics, and situational awareness. Petabyte databases and data gathering rates of tens of TB/day are already with us; a thousandfold increase may be envisioned well within the next decade.

1. UQ for approximate algorithms: One of the key challenges is developing very fast algorithms for data analysis, whether statistical in nature, or having to do with aspects such as graph and network analysis and pattern recognition. The severity of this problem is such that even O(NlogN) algorithms may be far too slow, and O(N) or even O(1) algorithms may be needed. However, the only available fast algorithms are likely to be approximate ones, in which case building a robust UQ infrastructure for the predictions from these algorithms is essential.
2. UQ and extreme modeling and simulation I: The datasets produced by extreme scale computing will soon be as rich as experimental or observational databases. Indeed, large future experiments and observational campaigns are already being designed in a feedback loop with simulations. This imposes a serious requirement on the validity of the modeling and simulation process which in turn requires a new class of UQ methods able to deal with problems such as high dimensionality and predictions for extreme values.
3. UQ and extreme modeling and simulation II: Very fine-grained and complex simulations have a natural application as test-beds for problems that cannot be handled experimentally, e.g., disaster response. In cases such as this, it is unlikely that any single computational model can be realistically accurate. However, the fact many intervention strategies (and many models) can be separately investigated immediately poses the need for a UQ paradigm able to assess the usefulness and risks of these strategies, given the underlying limits of the models and simulations.
4. UQ co-design with new computing architectures: It is widely accepted that supercomputing architectures will undergo major changes over the next decade. Next-generation applications must deal with severe concurrency and latency challenges and must be developed in close concert with the evolving architectures. This will also certainly be true of UQ methods and frameworks which must be flexible enough to encompass next-generation hybrid supercomputers, data-intensive supercomputers, very large cloud computing platforms, and special-purpose machines.

Summary of Research Direction

This research direction is driven by the proliferation of data both sensor-derived and digitally archived, as well as generated by large-scale computers. The scale of the data throughput and size renders, in many cases, answering classes of precise questions a meaningless exercise, even if the data is well-characterized. Thus, one task imposed by this new arena is to develop UQ strategies for inherently approximate analyses. Additionally, the scale of the data is such that it will contain (possibly) very high-dimensional dependencies, and dealing with them will require robust UQ-controlled techniques of model and data reduction, including controlling uncertainties from combining large-volume disparate data sources. Finally, the UQ methodology will be required to be sensitive to extreme scale computer architecture (both conventional and data-intensive) as this itself will evolve considerably.

Expected Computational and Scientific Outcomes

This research direction will be an essential aspect of information extraction from large datasets, especially as these datasets quickly scale beyond the reach of current analysis methodology. Many of the applications will be associated with major scientific and engineering efforts, national security data gathering and databases, and important social issues, all areas where a reliable and robust UQ paradigm should be considered a key, if not the dominant, requirement.

Potential Impact on National Security

The impact on national security is very significant. Decision makers will need to understand the ramifications of various actions -- undertaken both by them and others -- in an increasingly complex and data-rich environment. The ability to extract robust information and correlations, and to be able to quantify the uncertainties associated with certain actions will be a key aspect of UQ for national security. Additionally, at a lower level, many aspects of national security will require dealing directly with UQ issues for datasets at extreme scales – cyber security, counter-terrorism, disaster response, and situational awareness, are all obvious examples.

Combining Disparate Models and Data Sources

With many physical systems, there does not exist an integrated computational model – or code – that incorporates all relevant processes from which scientific inferences can be made. Typically there are a number of computational models available to model different aspects of the system. These separate models, though they may share some commonality, typically focus on different aspects of the system. Similarly, there are often a wide variety of data sources available to inform about a given physical system. A priority research direction is the development of a conceptual framework and methodology for making scientific inferences with the aid of these disparate models and data sources.

Scientific and Computational Challenges

There are a number of scientific investigations of importance to national security that motivate this need for combining disparate models and data sources. These include

- Inferring material behavior: computational models for materials are now available on multiple scales of resolution – from atomic, to meso, to macro scales. The atomistic models, though computationally demanding, are very nearly first principles models. Hence they can be used to help infer bulk properties of the material, such as equation of state or strength. The development of UQ methods are essential for combining these models to infer bulk material properties at temperature and pressure conditions that cannot be accessed in laboratory experiments, particularly for furnishing realistic model uncertainty estimates appropriate to such extrapolations.

- Estimating, tracking and managing greenhouse gas (GHG) fluxes: while no fully integrated earth-system model exists (or is likely to exist) that incorporates all processes relevant processes to atmospheric GHGs, there is a wide variety of models and data sources to aid in the estimation, tracking and management of GHG fluxes. Relevant models might include atmospheric transport, ecological dynamics, the carbon cycle, land use, social behavior and energy infrastructure. Relevant data sources are equally varied – econometric inventories and summaries, census information, land, sea and air based sensors, satellite observations, and isotopics are just some of the data sources. A framework and methodology is clearly needed to infer the current state of GHG fluxes. These are also needed to infer the impacts resulting from potential mitigation strategies.
- Inferring climate change at the regional level: while climate change predictions are typically made with large-scale global circulation models, planners need to know how such changes will affect their local climate. Will current water sources dry up? Will agriculture remain tenable? etc. Approaches for downscaling coarse level climate change information to the regional level exist, but uncertainty quantification for these regional predictions is not well developed. UQ methods for this application require that one combine global and regional models with local information such as weather, topography, groundcover and hydrology to predict the effect of climate change at this local level.

In addition to the specific application areas called out above, relevant scientific challenges exist in nonproliferation where a wide variety of information is available (from intelligence to sensor signals), and models may include agent-based or socio-technical simulations.

Summary of Research Direction

Just as research in "data fusion" focuses on the development of new frameworks, methods and algorithms for making use of diverse information sources for making inferences, this research direction extends this notion to incorporate the use of diverse computational models as well. While the construction of tightly coupled multi-physics codes tackles this problem head on by fusing multiple models into a single code, this solution is labor intensive and very specific to a particular application area. Also, management and upkeep for such codes can be a daunting task. This research direction proposes the development of new UQ methods to link different models and data sources into a common inferential framework.

This looser coupling of different models will necessarily lead to a greater reliance on physical observations. Hence, the ability to incorporate all available data and information sources – possibly in a dynamically updating situation – will be crucial. These sources may include physical observations, experiments and expert judgment. It is expected that new statistical approaches for meta-analysis and hierarchical modeling will be developed to make the required inferences. On the computational side, such inferences will undoubtedly require many model runs from each of the models involved with the analysis. It is also expected that approaches to determine what combination of model runs are required and how to optimize their allocation to high performance computing resources will also be developed.

Expected Computational and Scientific Outcomes
If successful, this line of research will greatly improve the ability of different data sources and computational models to make inferences about complicated physical, or social, systems. Such research will also give insight on how to sharpen, or improve the resulting inferences. For example, which models should be improved or altered? What will be in impact of including additional data sources or computational models?

Potential Impact on National Security
The payoff of such research will be an agile framework for putting together available and relevant models and data sources to answer important questions relating to national security. The ability to assess system response, with uncertainties – based on the available models and data – will be crucial for decision makers to formulate responses to potential threats to the national security.

UQ for Emergency Response
Future national security threats include events precipitated by weapons of mass destruction, natural disasters (including weather related calamities), man-made disasters, pandemic events that would have biological impacts, and potential threats to the population caused by human activities that could result in significant changes in the ecosystem. Effective, timely response to such threats requires accurate assessment of the scenario and its possible evolutionary state, coupled with useful estimates of the risks and consequences associated with potential response options. Accurate assessment of a given threat scenario or a particular response option can be an extreme computational challenge by itself. However, to be of the most use to emergency responders, such an assessment must be timely and must include an estimate of the uncertainties associated with the various possible evolutionary states and potential response options. Quantifying uncertainties for emergency response, therefore, adds another dimension to the requirements for exascale computing.

Scientific and Computational Challenges
Predictive modeling of disaster scenarios and response options requires advancements in physics models and computational capabilities. The challenges associated with the physics modeling capabilities are similar to those discussed elsewhere in this report. The challenges associated with the uncertainty quantification aspects of emergency response are related to the near real time decision support requirement for effective response. One challenge is the integration the scientific modeling capability with the historical and prompt data describing the scenario. Such data includes the following information:

- Source term (nuclear, chemical, biological, environmental)
- Emplacement environment (buildings, structures)
- Geographical terrain and features
- Local and regional weather conditions at time of release
- Population density and distribution
- Available response and mitigation options

The challenge for uncertainty quantification is to estimate the variability in the consequences of the response options from the potential variability of model predictions based on the uncertainties in the scenario information. The large number of variables associated with each of the inputs as well as the coupling between the associated models needed to determine consequences in real-time requires a very large number of calculations at large scale.

Summary of Research Direction

There is a spectrum of possible approaches for developing the capability for emergency response decision support based on physical modeling with quantified uncertainties. The spectrum runs from the exclusive use of pre-computed databases of scenario and response options and the associated uncertainties to near real time modeling and uncertainty quantification on demand. The pre-computed database end of the spectrum suffers from the obvious drawback of not accurately reflecting a particular threat and currently available response options, thereby adding uncertainty to those previously estimated. The near real time modeling approach requires more computing capability than will be practically available, even at the exascale. Therefore, a practical capability will require some combination of the two approaches. The proposed approach is to couple databases of historical data and high fidelity, pre-computed scenarios, response options, and the associated uncertainties with current information to produce accurate, scenario-specific support for emergency response decisions.

Therefore, the focus will be on the co-development of pre-computing strategies with novel analysis techniques that can efficiently exploit existing information databases (e.g. maps, terrain, buildings), prompt data (such as weather conditions and availability of emergency services), and computational model runs to give critical information regarding the scene and to evaluate and propose potential responses. Particular activities might include:

- The establishment of detailed worldwide databases of terrain, buildings, and population distribution for baseline calculations based on satellite data mining
- Develop a capability to couple weather or seismic data in real-time to address structural impacts and/or dispersion events anywhere in the world
- The creation a comprehensive database of source term calculations under a variety of emplacement geometries to predict patterns of dispersal and damage for weapons of mass destruction
- Establishing a surrogate modeling capability for interpolation between pre-computed, high fidelity physics models and propagation of uncertainties
- The development of methods that use simulation models of mass-evacuation strategies to optimize moving personnel to safety in a crisis
- The development of approaches to estimate structural response for a variety of classes of buildings and use satellite data to populate those classes

Expected Computational and Scientific Outcomes

The primary product of this PRD will be the development of capabilities necessary to support an accurate, timely emergency response decision support system. Such capabilities include those listed above and might include additional products such as:

- Development of satellite photo data mining techniques to construct local structural, traffic, and population models of target areas
- Development of a seamless ability to couple real-time weather to the geographic database to calculate dispersion and effects

These capabilities could be integrated into a nationwide decision support system that combines laptop-based software for emergency responders coupled with remote database capabilities and computational resources.

Potential Impact on National Security

Risk-based emergency response decision support will enable emergency responders to optimize resources and minimize consequences of natural and manmade disasters. This capability also provides accurate information to the national security infrastructure for assessments of impacts on national security and appropriate marshalling of national security assets.

References

Brooks, H.E. and Doswell, C.A., 1996. A comparison of measures oriented and distributions-oriented approaches to forecast verification. *Weather and Forecasting*, 11, 288--303

Charney, J.G., 1979: Carbon Dioxide and Climate: A Scientific Assessment. National Academy of Sciences, Washington, D.C., 22

R. S. Lindzen and Y.-S. Choi, On the determination of climate feedbacks from ERBE data, *Geophysical Research Letters*, VOL. 36, L16705, doi:10.1029/2009GL039628, 2009

B. J. McAvaney, C. Covey, S. Joussaume, V. Kattsov, A. Kitoh, W. Ogana, A. J. Pitman, A. J. Weaver, R. A. Wood and Z.-C. Zhao, 2001: Model Evaluation. In: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Meehl, G.A., T.F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M. Gregory, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver and Z.-C. Zhao, 2007: Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M.

Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

B. D. Santer, T. M. L. Wigley, M. E. Schlesinger and J. F. B. Mitchell, 1990: Developing Climate Scenarios from Equilibrium GCM Results, Report No. 47, Max Planck Institute for Meteorology, Hamburg.